BioClust User's Guide

by Serguei Sokol (sokol(a)insa-toulouse.fr)

First release 24/05/2004
Last updated 07/09/2007

platform “Biopuce” INSA/DGBA
135, av. de Rangueil,
31077 Toulouse cedex
France
https://biopuce.insa-toulouse.fr

Copyright (C) 2004, platform “Biopuce”, Toulouse Genopole. Permission is granted to freely print, copy, translate or distribute this document for educational purposes, provided this copyright notice is preserved.

The original and most upto date copy of this document can be found on the web page http://biopuce.insa-toulouse.fr/ExperimentExplorer/doc/

Contents

1 BioClust Form Items. **4**
1.1 Analysis tab. 4
1.2 Clustering tab. 6
1.3 Variable/Norm. tab. 8
1.4 Gene Selection tab 11
1.5 Category selection tab 13
1.6 Stats tab 14
1.7 An. profiles tab. 16

Introduction

BioClust is a web service hosted by platform "Biopuce" which is a part of Toulouse Genopole and is located in INSA/DGBA. This service is aimed to help platform users to compare transcriptome data from multiple biological conditions and select significantly changed genes according to a desired expression pattern. BioClust shares with BioPlot same data treatment methods such like normalization, Student's test or gene selection. For a user starting to work with BioClust, it may be useful to start by BioPlot User's Guide (`http://biopuce.insa-toulouse.fr/ExperimentExplorer/doc/BioPlot.pdf`) which presents some theoretical considerations about transcriptome experiment.

BioClust does not provide any evolved graphical tool for clustering manipulation but it can be very helpful to prepare data which can be submitted to a software specialized on clustering or other multivariate statistical methods, see, for example, Tigr Multi-Experiment Viewer (tmev) on `http://www.tigr.org`.

The rest of this guide is a detailed description of form fields which may be found in BioClust.

The reader is supposed to be familiar with theories and techniques employed in transcriptome experiments as well as with some advanced statistical techniques like clustering and Linear Discriminant Analysis (LDA).

Chapter 1

BioClust Form Items.

In this chapter, we describe the parameters that user can tune in BioClust to analyze transcriptome data corresponding to two or more conditions referred as X_i, where i is condition number. To analyze only two biological condition, BioPlot provide reacher analysis environment including scatterplot, zoom on image and link to external genome databases.

BioClust is a web service on `http://biopuce.insa-toulouse.fr`. Users have personal accounts and protected access to their data. When user connects and chooses BioClust in tool frame at left hand side, a BioClust form is opened in main frame. The names and comments in the form are intended to be self explanatory provided that user is familiar with transcriptome data treatment methods. The form is organized in following tabs :

Analysis where analysis to treat can be selected and distributed among columns, one biological condition per column.

Clustering where user makes his choice on data treatment method and on kind of result presentation.

Variable/Norm. where user chooses a quantification variable and selects options on background correction, normalization and log transformation.

Gene Selection regroups options relative to filtering data.

Cat. Selection Category selection is intended for gene filtering by ad hoc functional categories and by categories defined in Gene Ontologies (GO).

Stats has fields concerning statistical data treatment.

BioClust has not extended graphical possibilities. If you need an elaborated graph, you have to export data from BioClust to some software having such possibilities.

In following sections we review all form fields.

1.1 Analysis tab.

Group name for LDA (text) This entry is for labeling groups of columns. It should be filled only if you want to proceed Linear Discriminant Analysis (LDA) treating columns like individuals and rows (gene expressions) like observed variables. Usually for LDA, you have to put one analysis per column and to label each column by a group label. You need to define two or more groups each of which is composed of at least two columns. Thus columns belonging to the same group will have the same label.

The purpose of LDA is to find linear combination of observed variables (here gene expressions) which have the most discriminant power for submitted groups. LDA maximizes Fisher statistic, i.e. the ratio "inter-group variance over intra-group variance". The variances are calculated according to some linear combination of observed variable. LDA gives up a sorted list of genes according to their weight in linear combination which can be interpreted as importance or contribution to the discrimination.

The main LDA result is a sorted table of genes. The ordering is done according to the contribution of a given gene to the first discriminant variable (LD1) which maximize the ratio mentioned above. Thus, the most discriminant genes are in the top of the table. Absolute values of values reported in the column LD1 can be considered as contribution weights of a given gene to LD1.

We report a Fisher statistics (i.e. ratio inter-/intra-variance) as well as corresponding P-value for every discriminant variable. Naturally, it is the first variable (LD1) which is the most important for us. For example, if P-value corresponding to LD1 is low (say < 0.05) than LD1 is a good discriminator thus the most contributing genes to LD1 are, in their turn, good discriminators of biological diversity.

For more information on LDA see, for example, `http://www.isip.msstate.edu/publications/reports/isip_internal/1998/linear_discrim_analysis/lda_theory_v1.1.pdf` or, in French, `http://www.lsp.ups-tlse.fr/Besse/pub/sdm2.pdf` (look for "Analyse Factorielle Discriminante").

1-st column (multiselect menu) In this menu user chooses biological conditions which may be interpreted as control conditions depending on the choice in BioClust > Clustering > Table content (see description of this item here after for more details).

After you have clicked on "put in 1-st column", this multimenu will become "2-nd column" and so on, up to a desired column number.

When an option "BioClust > Clustering > Table content" (described here after) is set to "Ratio on coupled analysis (channel)", analysis chosen in this menu are treated as test analysis, i.e. the ratios are calculated as intensities from selected analysis over the intensities coming from complementary channel on the same slide. The last are detected automatically by BioClust.

matching (button + text field) click on this button to make the menu contain only analysis whose names are matching the content of the neighbor text field. This is a practical mean to reduce analysis choice. This feature is particularly useful when

the number of analysis become very important. Note that on recent browsers (Netscape 7+ or InternetExplorer 6+) the menu content becomes conform to the text field at each new entered letter, such that you don't need additional click on "matching" button.

back to #-th column (button) This button is visible only from the second column selection and allows to return to the selection of precedent column.

Selected (integer) This is a recall of how many arrays are already selected for current column.

put in 1-st column (button) Click on this button when you have finished the selection of analysis for the first column. You will be presented almost identical Analysis tab. The multimenu for analysis selection will take name "2-nd column", this button will become "put in 2-nd column" and a button "back to 1-st column" will appear. So you will be able to select analysis for second, third and so on column.

Finish (button) This button is visible in all tabs. It should be clicked only after have set all parameters in all tabs. To pass to other tabs, click on the corresponding tab name.

1.2 Clustering tab.

Gene ID (menu) Historically, the gene naming is not something unique and standard. So various choices are possible. In this description we use the following terms for gene names : short name (like *ACT1* for actin), systematic name (like *YFL039c* for yeast actin[1]), full name (like *actin*), user's code (whatever user has used to identify his spotted material). The options of this menu are:

Default short name if exists, systematic name otherwise.

Short name only short name is used. If it is void than gene ID will be void in resulting list.

Systematic name only systematic name is used as gene ID.

Sys. name; short name both names are used and they are separated by semi-column and a space.

Full gene name is self explanatory.

User's code idem.

One check box option can modify any choice made in menu:

append well will append the plate coords of well containing the gene, e.g. *ACT1; 1A12*

append user's code idem for user's code.

[1] For other organisms, GeneBank entries are often used as systematic names.

append fullname idem for full gene name.

Table content (menu) Options are

- Ratios on first column. This option is equivalent to membrane mode in BioPlot. An average value of gene intensities is used to calculate the final ratio of each column over the first one.
- Ratios on coupled analysis (channel). This option is equivalent to slide mode in BioPlot. A ratio is calculated for each spot for all couples of analysis. A couple is made of analysis coming from the same slide. The final ratio appearing in columns are the two-stage average of spot ratios (first stage: over spot replicates, second stage: over slides).
- Expressions: Up=1; Equal=0; Down=-1. According to this option, the resulting table will be composed of expression labels: "1", "0" and "-1". The decision about over- or under-expression of each gene is done following the selection in "BioClust > Stats > Expression changes based on".
- Variable values. This option makes the result table contain the intensities, not ratios. The variable quantifying spot intensity may be chosen in "BioClust > Variable/Norm. > Variable to use".

append P-value (checkbox) if checked, this option add columns with P-values to columns requested as "Table content". Columns with P-values are alternated with others columns such that a gene has its ratio and corresponding P-value for a given condition side by side.

Hide first column (checkbox: Yes). Check this box if the selected option in previous menu is "Ratios on first column". Thus, the first non informative column having only "1" or "0" (depending on the option "Variable/Norm. > Take log10") is not shown.

Result type (menu) Options are:

- html table with colors. According to this option, the cells of resulting html table will be colored in red or green depending on over- or under-expression of a given gene in a given biological condition. If we cannot test the expression change (because of data lack, for example), the corresponding cell will have gray background.
- html table. The same table but without colors in cells.
- plain text. This option is useful for exporting data from BioClust to any other software.
- h-clust by R:hclust{mva}. The result will be an image with a dendrogram corresponding to hierarchical clustering. This clustering is done by a function hclust from the library mva which is part of R script language (see `http://www.r-project.org/` or `http://www.lsp.ups-tlse.fr/`

Besse/pub/TP/r/tpintro.ps for a brief introduction in French). R functions may need some number of parameter settings which are not accessible from BioClust. The resulting dendrogram corresponds to default options, i.e. Euclidean distance and average link. Here, the dendrogram is provided only for quick visual analysis and is limited to 500 genes. A thorough study should be conducted in R or any other statistically oriented software like tmev on www.tigr.org. For more information on hclust see help(hclust) in R. For an introduction in clustering, see, for example, http://www.statsoftinc.com/textbook/stcluan.html or http://www.lsp.ups-tlse.fr/Carlier/.Hyper/polyclass/node1.html (in French).

- heatmap by R:heatmap{mva}. Hierarchical clustering is performed both on genes and conditions like in previous item. The results are presented in so called "heatmap" form where each table cell is coded is some color (green for low values, red for high values). Such presentation facilitates a visual detection of dissimilarities between genes and condition. This can be useful for a subjective quality control.
- Lin. Discr. An. by R:lda{MASS}. This option defines a result type which is provided by a lda function from library MASS available in R mentioned above. The purpose of Linear Discriminant Analysis is briefly presented in the description of the field "BioClust > Analysis > Group name for LDA" on page 4.
- Scatterplot pairs. This kind of presentation may be useful for quality control. User can select a set of arrays, one by column, and see the scatterplots of all possible column pairs. The values reported on scatterplots depend on the choice in menu "Table content", upper in this tab. For exemple, if "Table content > Ratio on coupled analysis (channel)" is chosen then (log) ratios on each slide will be compared with (log) ratios of all other selected slides. Correlations, anti-correlations or other kinds of point clouds observed on such scatterplots may be relevant for potential problems of labeling, scanning and so on.

1.3 Variable/Norm. tab.

Variable to use (menu) This menu gives the list of variable to quantify spots. Most used variables are "Mean Intensity" and "Median intensity". Mean and median are calculated over pixels of a spot by image analysis software. These and others spot measures are imported in data base from text files generated after image analysis. As various applications are used for spot detection and quantification and they measures their own statistics, not all choices are meaningful for all analysis. The full list of options in this menu is :

- Variable Selected during Image Analysis
- Mean Intensity

- Median Intensity
- Weighted Mean Intensity
- Mean Gaussian Intensity
- Median Gaussian Intensity
- Mean Statistical Intensity
- Median Statistical Intensity
- Most frequent pixel intensity
- Central Intensity
- Maximal Intensity
- Minimal Intensity
- Sum Intensity
- Background level
- Background Median
- Background Mean
- Spot Standart Deviation
- Spot Variance
- Background Standart Deviation
- Total Background
- % pixels superior to background
- % pixels superior to 1.5*background
- % pixels superior to 2*background
- % gaussian pixels superior to background
- % gaussian pixels superior to background
- % gaussian pixels superior to background
- % statistical pixels superior to background
- % statistical pixels superior to 1.5*background
- % statistical pixels superior to 2*background
- % pixels superior to Bg+SD
- % pixels superior to Bg+2*SD
- % pixels saturated

"Variable Selected during Image Analysis" is particular. Some image analysis applications offer normalization features or other data treatments. For this reason, this variable can not be corrected by background. This option is deprecated and will be removed in the future.

Normalization type (menu) Choices are:

- no normalization
- housekeeping genes/spike control
- stable majority (by histogram)
- all spot's mean
- all non negative mean
- lowess
- quantile

Use spots marked "Reference" as control genes (checkbox: Yes) This option used only when "housekeeping genes/spike control" is selected in previous menu. If checked, the spots having attribute "Reference" set, are used to calculate normalization coefficient. To mark spots as references, see the BioPlot User's Guide for instructions which can be found in TabView description. Checking this box puts the option "housekeeping genes/spike control" as selected in previous menu.

Control genes coords (text) This option used only when "housekeeping genes/spike control" is selected in "Normalization type" menu and previous checkbox is not checked. This field can contain a comma separated list of well or spot coords corresponding to genes that should be used as references in normalization. Well or plate coords are given as <plate nb>[<letter>[<number>]], e.g. `1A14` means plate 1, well A14. <letter> and <number> are optional. If number is omitted, the whole line is used. If <letter> and <number> are omitted, the whole plate is selected. It is possible to define a rectangular region on a plate by introducing the left upper and right low conner well coords. For example, `2C3:D12` defines a region on the plate 2 containing wells in rows C to D and between columns from 3 to 12.

Spot or array coords are of the form <row><space><column>, where <row> and <column> are integers starting at 1. For instance, `11 23` defines a spot at row 11 and column 23. All arrays in our data base are oriented to have a spot corresponding to the well 1A1 in upper left corner. This orientation is 90° rotated compared to vertically oriented slide scans. Usually, on vertical slide images, the well 1A1 has its spot in upper right coin. You can define a rectangular region by defining the starts and ends for rows and columns as follows : [<row start>]:[<row end>]<space>[<col start>]:[<col end>]. For example, `12:15 23:40` defines an array region having rows between 12 and 15 and columns between 23 and 40. Any of region border (start or end) is optional in which case the limit value (1 for start, maximum for end) is taken. Thus `12: :40` defines a region with rows from 12 to the max row and columns from 1 to 40.

Apply post normalization factor (checkbox: Yes) If checked, all spots in every analysis are multiplied by a user defined factor. This is used in very particular cases. Don't check this box, if you don't need to correct normalized data by hand. This factor (one per analysis) can be defined on the "Browse analysis" page accessible from a tool frame on the left of user's web space.

Subtract background (menu) Apply or not background correction. Options are

- no
- local background
- average of negative spots
- local+negatives corrected by their bg

Last two options may be useful for membranes with clones grown on them and having DNA sequences inserted in their own DNA. Such array technique can lead to a situation where clones with a void insert have a signal above the background. These negatives control should have signal as close to zero as possible, so it is more appropriate to correct by the mean of negative controls than by local background. In other situations, "local background" should be good choice.

Use average of spots marked "Negative" as bg level to subtract (checkbox: Yes) This checkbox is taken into account only when one of the last two options of precedent menu is chosen. For marking spots as negative, see BioPlot User's Guide. When this option is checked, it puts "average of negative spots" as selected in precedent menu.

Negative gene coords (text) If "average of negative spots" or "local+negatives corrected by their bg" options of "Subtract background" menu are selected, this field can define the spots that should be used as negative controls. User can enter a comma separated list of well or array coords following the same coord conventions as in "Control gene coords" field on this tab.

Take log10 (radiobox) Choices are "No", "Yes" and "Yes, but final results are in original scale". If the last option is chosen, the log transformation is applied and all statistical treatment is done on log-transformed values. Only final results will be back transformed from log to original scale for easier interpretation of ratio values. It is highly recommended to apply log transformation to bring better statistical properties to data passed through statistical tests.

Lowest value limit (real number) It may happen after the background subtraction that a spot intensity become zero or negative. This is not desirable, if we want to apply log transformation. So to prevent too low values, they remain at lowest allowed value defined in this field. Don't confound this feature with cut off by intensity. No spot is cut here. Conversely, due to this feature, low spots are preserved in final list. To disable such preserving, cancel any content of this field (don't leave "0").

1.4 Gene Selection tab

Filters defined in different fields of this tab work like intersections, i.e. if you select spots coming from first plate and the genes whose name start by "a", you obtain genes starting by "a" coming only from the first plate. Conversely, in the same field, your

selection is considered like union, i.e. if you select first and second plate to put on the result table, than genes coming from these two plates will be shown. An intersection in this case would be meaningless as the intersection of first and second plate is void.

When applicable, any selection may become exclusion if preceded by tilde "~".

By plate(s) region (text) You can enter a comma separated list of plate coords using the same coord convention that for the field "Variable/Norm. > Control gene coords".

By membrane(s)/slide(s) region (text) Use a comma separated list to define regions of interest on your arrays. The coord conventions are the same as for the field "Variable/Norm. > Control gene coords".

By name (text) You can select one or more genes by their names. The names entered in this field should be in accordance with the choice in "Analysis > Gene ID" menu. The names can be separated by comma, by tabulation or by new line character. It is perfectly possible to copy and past a gene list from an other application such as a spreadsheet.

The gene name can have a wild character "*" which replaces any sequence of characters. Gene names are not case sensitive. For example, `y*` selects all genes starting by "y" or "Y".

The wild character alone has a particular sense. If entered, it selects all spots corresponding to some gene id. Spots don't having gene identity are excluded from selection.

By stored gene list (text) User can give a file name with stored gene list. This will select the genes of this list. User can store a gene list both in BioPlot and BioClust. At the end of result table, there is a form to fulfill, to store currently shown gene list. A button "..." can be used for help in file choosing.

By (X_{ref}+X_i)/2 cutoff when (X_{ref}+X_i)/2 is (menu, number and other menu) This is excluding field. Here you define the spots to exclude from from results. Very low spots may be annoying to see in the final results. They can exhibit very high ratios but they are not very reliable. When some real number is entered in text field, the result of exclusion can be read as, for example, "cutoff when (X_{ref}+X_i)/2 is less or equal to <some number> in all columns". Another choice in the first menu is "greater or equal" which is complementary of the option cited in the example and can be useful to visualize the low spots that are cut. The second menu has an option "in at least one column" and mean that a spot will be excluded from the result if it is too low simultaneously in control and any (even only one) test condition.

By expression pattern (menus) All menus (one per defined column) have the same options relative to expression decision:

- All
- Up

- Norm.
- Down

Setting these menus to desirable values, user can extract one or several expression profiles.

By intensity. Put spots only if #% pixels > bground (real number, checkbox) If checked, spots having the percentage of pixels above the background lower than defined threshold, are excluded from scatterplot and from result table. The statistics "percentage of pixels above the background" is measured and provided by image analysis application. This option is not very reliable to eliminate low spots. It is preferable to use "By $(X_{ref}+X_i)/2$" field.

By expression (checkboxes) This is another excluding field. The choices are

- Exclude over-expressed genes
- Exclude normally expressed genes
- Exclude under-expressed genes
- Exclude unknownly-expressed genes
- Exclude unknownly/normally -expressed genes

The decision if a gene is over or under expressed is related to the field "Stats > Expression changes based on". Note that if you exclude, for example, normally expressed gene, then only genes normally expressed in all conditions will be excluded. If a gene has changed expression in at least one condition, it will be preserved in results.

By type (menu) Options of this menu are

- all
- reference
- negative
- not negative

This can be useful to visualize the spots/genes corresponding to any chosen type. To see reference genes may be of particular interest if the normalization by "stable majority" is used. In this case, a users don't know a priori what are the genes that are selected as stable majority. So this option lets visualize them.

1.5 Category selection tab

Features of this tab are identical to those of BioPlot > Cat. Selection.

Category selection or for sake of brevity cat.selection, is composed of four fields - one for gene filtering by ad hoc functional categories and three others for filtering by Gene Ontologies (GO). To use these options, user has to provide an annotation information for his organism of interest. The platform data base administrator has to treat and add this information prior to data analysis.

By functional category (text and select button) user can provide a list of category names separated by ";" or at his convenience he can use a select window which opens after clicking on "..." button. User has to use ";" and not a coma "," to separate categories because some categories have a coma in their names.Categories are organized in a hierarchical tree-like structure. The tree branches are rolled or unrolled by clicking on a blue triangle. By checking a box corresponding to a functional category, user select all genes of this category and its subcategories. The name of the selected category is automatically added to the text field. Unchecking a category removes its name from the text field. User is allowed to use a meta-character "*" to replace any portion of category name when entering the name in the text field.

By GO biological process (text and select button)

By GO cellular component (text and select button)

By GO molecular function (text and select button) All three fields are functioning in similar way to functional categories. The main difference is that gene ontologies are defined by GO consortium in the same way for all organisms. So, there is no need to select an organism before selecting a category as it is the case for ad hoc functional categories.

Maximum hierarchical level to show (menu) As its name indicates, this menu controls the level of tree expanding when showing the category hierarchy both in select window and in tabulated results when ordering by a functional category or a gene ontology.

1.6 Stats tab

Spot aggregate function (menu) Options are :

- average
- bi-weighted mean

Duplicate aggregation may be done in classic way, by simple average or in more elaborated way, using bi-weighted mean. This last method is meaningful only when the replicate number is relatively high, e.g. 11, like on Affymetrix slides.

Spot aggregate (menu) Options of this menu are:

- by well
- by gene name

The most currently used option is "by well". However, "by gene name" can be useful if plate plan of compared analysis are different. Another potential application of this option is for eliminating of multiple entries in result table if some genes are present in more than one well. This option is mandatory if "BioClust > Clustering > Result type" is set to "h-clust by R:hclust{mva}" as hclust doesn't support neither repeated nor unnamed entries.

Expression changes based on (menu) This option is without effect if no checkbox is checked in "BioClust > Gene Selection > By expression". Choices in the menu are

- ratio thresholds
- Student's test
- ratio AND Student's test

The first option "ratio thresholds" is intuitive and seems simple for biologists. Unfortunately, this method can result in a high number of false positives. It is highly recommended to combine this filter with Student's test.
When "Student's test" is selected, user is exposed to multiple tests problem discussed in BioPlot User's Guide. So the combination of ratio thresholds and Student's test is probably the most appropriate choice.
To be able to apply Student's test, user has to supply at least two independent array repetitions. However, the power of statistical test on only two repetitions may be deceptive. You can estimate the number of repetitions on `http://biopuce.insa-toulouse.fr/microarray/exp_numb.php`.

Overexpression threshold (> 0) (positive real number) This field is used when one of checkboxes "BioClust > Spot Selection > By expression" is checked and an option selected in the precedent menu concerns ratio.

Uderexpression threshold (> 0) (positive real number) Same remarks.

P threshold in Student's test (0 < P < 1) (positive real number less than 1) This field is used when one of checkboxes "BioClust > Spot Selection > By expression" is checked and an option selected in the menu "Expression changes based on" in this tab concerns Student's test.

Error estimate without gene-dye interaction (radio check boxes) Three options of this field can be used to choose the way to treat an error estimate in dye-switch experiment:

- "No" (default) corresponds to classical Student's test with classical error estimate for all genes.

- "Yes for all genes" means that error will be estimated without gene-dye effect contribution for all genes.
- "Yes for genes giving lower P-value" corresponds to automatic choice between two kind of error estimate. The error estimate canceling gene-dye effect is used only if it gives better P-value, otherwise a classical Student's test is used. This treatment avoid the loss of statistical power for genes not involved in gene-dye interaction.

Sorting criterion (menu) User can order data according to one of the following criterion:

- Expression
- Well's position
- Membrane position (by line)
- Membrane position (by column)
- Gene id
- Functional category
- GO biological process
- GO cellular component
- GO molecular function

Note that "Expression" in this menu means a discrete variable taking three values 1, 0 and -1 depending on expression change test. Ordering by Expression will form gene groups having the same expression profile. These groups are easily detectable in html page if colors are used to highlight expression changes.

Ordering (menu) Options are:

- ascending
- descending

1.7 An. profiles tab.

Save current analysis parameters as (text) file name in which it will be stored the current set of analysis parameters such like type of normalization, filtering rules and so on. A set of analysis parameters is called *analysis profile* and storing it in a file can be useful to gain an effort when a user has to analyze an experiment series in same conditions. It diminishes also an error risk in choosing parameters. Analysis profile files are stored in `anprof.bioclust` subdirectory of user's space. To choose an existent file or a subdirectory, user can click on a button with ellipsis .... All files have `.kvh` extension. If user omits this extension in file name, it will be added automatically.
There are two predefined profiles :

predefined/mbr_normtot_exp.kvh and
predefined/slide_normtot_expr.kvh

for membrane and slide mode of averaging respectively. These profiles enable

- use of pixel mean as spot intensity
- local background correction
- log-transformation of intensities
- normalization by total mean of all spots
- averaging of spot replicates by well
- expression decision based on Student's test and ratio thresholds
- no filtering rules
- sorting by expression (taking values in {1; 0; -1; NA}) in descending order.

Predefined parameters can be a start point for defining user's own profiles. They cannot be overwritten.
When a file name is chosen, user has to click on "Save" button to actually write the file.

Read analysis parameters from (text) file name from which an analysis profile will be read.

Delete analysis parameters (text) file name or a void subdirectory to be deleted.

Troubleshooting.

Warning messages.

When user makes some inconsistent choice, the result page can contain one or more warning messages which explicate the inconsistency. Normally, after reading the warning message in a particular context, it becomes clear what selection should be changed to remove inconsistency.

Other problems.

In case of:

- the warning message is not sufficiently explicit;
- there is problem with BioClust that you cannot resolve by yourself;
- you would like to request a particular feature to be added to BioClust;
- you want to report a bug in BioClust;
- there is a correction to do on this documentation;

you can contact Serguei Sokol by email `sokol(a)insa-toulouse.fr` or by phone +33 (0) 561.55.96.87

In case of a problem with your analysis, it can be helpful to include in your email a copy of a link under the result table in BioClust (if any). You can click with right button on this link a choose "Copy link location" from contextual menu and past it in email composing window.

Credits

The work on BioClust and on documentation has benefited from various financial supports from INRA and CRITT-Bioindustrie. Platform director Jean Marie François and platform technical director Véronique Le Berre helped the author to understand transcriptome experiments and by fruitful discussions contributed to this documentation.